(19) BUNDESREPUBLIK DEUTSCHLAND

DEUTSCHES PATENT- UND MARKENAMT

(12) **Offenlegungsschrift**

(10) **DE 100 50 808 A 1**

(51) Int. Cl.[7]:
G 10 L 15/22
G 10 L 11/00
H 04 M 1/66
H 04 M 3/50

(21) Aktenzeichen: 100 50 808.1
(22) Anmeldetag: 13. 10. 2000
(43) Offenlegungstag: 16. 5. 2002

(71) Anmelder:
VoiceCom AG, 90449 Nürnberg, DE

(74) Vertreter:
Leonhard Olgemöller Fricke, 80331 München

(72) Erfinder:
Schimmer, Klaus, 90513 Zirndorf, DE;
Plankensteiner, Peter, 91054 Erlangen, DE; Harbeck, Stefan, 91052 Erlangen, DE

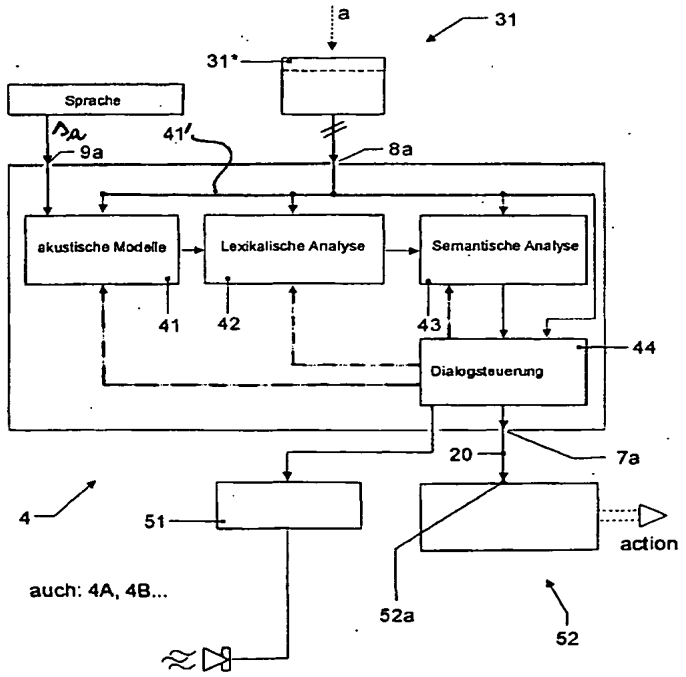
(56) Entgegenhaltungen:
DD 2 18 969 A1
US 60 88 669

Die folgenden Angaben sind den vom Anmelder eingereichten Unterlagen entnommen

Prüfungsantrag gem. § 44 PatG ist gestellt

(54) Sprachgeführte Gerätesteuerung mit Benutzeroptimierung

(57) Die Erfindung bezieht sich auf ein Verfahren zum Vorbereiten, Betreiben oder Anpassen einer sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technischen Gerätes (52), wobei ein Audiosignal (s_a) aus zumindest einem von einem Sprecher abgegebenen Wort einem ersten Signaleingang (10) einer Authentifizierungseinrichtung (2) zugeführt wird. Eine Sprechererkennung erfolgt aufgrund eines Authentifizierungsversuchs (2), um einen Sprecher als Individium oder eine objektivierte Gruppe von Sprechern festzulegen, der der Sprecher durch objektivierte Kriterien des auf ihn zurückzuführenden Audiosignals zuzuordnen ist, und ein entsprechendes Ausgangssignal (a, a*) abzugeben. Es wird ein zum festgelegten Sprecher oder der objektivierten Gruppe korrespondierendes Profil (33) aus einer Vielzahl gespeicherter Profile (32, P_i) ausgewählt (2, 31), gestützt auf das Ausgangssignal (a, a*) der Authentifizierung und das ausgewählte Profil (33) in eine Erkennungsumgebung (4) eingebunden oder geladen, um die Erkennungsumgebung auf den festgelegten Sprecher bzw. die objektivierte Gruppe hin anzupassen. Jedes der gespeicherten Profile (P_i) und das eingebundene oder geladene Profil (33) enthält Parameter, zur Beeinflussung zumindest einer in der Erkennungsumgebung (4) vorgesehenen Wortfolgeerkennung (42).



DE 100 50 808 A 1

DE 100 50 808 A 1

Beschreibung

[0001] Die Erfindung befaßt sich mit einem Verfahren zum Betreiben einer Steuerungseinrichtung, die sprachgesteuert ist. Die Erfindung befaßt sich auch mit einem Verfahren zur Vorbereitung, also zur Einrichtung einer solchen Steuerungseinrichtung, um sprachgesteuert arbeiten zu können. Ebenfalls betroffen ist ein Verfahren zur Anpassung der Steuerungseinrichtung, um an unterschiedliche Sprachsignale (Audiosignale) besser angepaßt zu werden. Sinngemäß ist das Betreiben übergeordnet, umfaßt also sowohl die Vorbereitung, wie auch die Anpassung während eines Betriebes und umschreibt als solches den Betrieb einer sprachgesteuerten Steuerungseinrichtung, allerdings sind Vorbereitung und Anpassung einzelne Funktionen oder Betriebsweisen innerhalb der globalen Gesamtbetriebsweise. Insoweit ist das Verfahren gemäß Anspruch 35 ein Betriebsverfahren zum Betrieb eines Gerätes, das mit Sprache (mit einem Audiosignal) steuerbar ist. Auch eine Einrichtung zum Ausführen der Verfahren ist vorgesehen (Anspruch 34).

[0002] Vorgelagert werden soll eine begriffliche Festlegung, um das Verständnis zu erleichtern. Soweit von einem Audiosignal oder einem Sprachsignal die Rede ist, ist dieses Sprachsignal nicht zwingend ein unmittelbar gesprochenes Wort, das sich nach Aufnahme über eine Erfassungseinrichtung, wie ein Mikrophon, als elektrisches Signal darstellt. Ebenso sind Sprachsignale auch off-line zur Steuerung einsetzbar, wenn sie als eine Datei zur Verfügung gestellt werden oder über Netzwerke zunächst übermittelt werden, bevor sie zur Steuerung verwendet werden. Das Sprachsignal im weiterhin verwendeten Sinne umfaßt also nicht nur die unmittelbare Sprache, sondern generell das aus ihr in irgendeiner Weise folgende Signal, auch nach einer Zwischenspeicherung oder einer Zwischenübertragung. So gesehen kann das Sprachsignal als zeitabhängiges Signal vorliegen, wie auch als ein Frequenzspektrum. Es enthält eine Information, die der Sprecher, also der Autor des Sprachsignals, vermitteln will. Dieser erste Bedeutungsgehalt des Audiosignals soll erfaßt werden und soll dem Gerät zugeordnet werden, um eine Funktion dieses Gerätes als "Action" anzusprechen.

[0003] Der Sprecher ist der Urheber des Audiosignals. Er muß dem Erkennungssystem nicht bekannt sein, er muß auch dem Vorbereitungsverfahren, Betriebsverfahren oder Anpassungsverfahren nicht unbedingt bekannt sein. Das System bereitet dann die Erkennung des Bedeutungsgehaltes vor, indem eine Vorbereitung durchgeführt wird.

[0004] ist der Sprecher bekannt oder zumindest hinsichtlich objektiver Kriterien im Rahmen einer Gruppe (im Sinne eines objektivierten Kreises von Personen) eingrenzbar, so kann das System eine eingeschränkte Anpassung erfahren. Der Betrieb findet dann mit dem angepaßten System oder mit dem vorbereiteten System statt. ist der Sprecher genauer konkretisierbar, liegt er als individuelle Person bereits in der Systemspeicherung vor, so kann das System unmittelbar auf ihn angepaßt werden.

[0005] Soweit ein "Sprecher" genannt wird, ist eigentlich sein Sprachsignal oder das Audiosignal gemeint. Diese beiden Begriffe werden synonym verwendet, das System arbeitet aber nicht mit einer optischen Erkennung des Sprechers, sondern mit einer Erkennung der akustischen Signale, die auf den Sprecher als Urheber zurückzuführen sind.

[0006] Aus dem Stand der Technik sind Telefonanlagen mit Sprachsteuerung bekannt, von denen einige erläutert werden sollen. Aus US 5,917,891 (Will) wird vorgeschlagen, das Anrufverhalten eines Benutzers zu verwenden, das aus einer Historie seines Anrufverhaltens hergeleitet ist und im Rahmen eines neuronalen Netzwerkes ergänzende Verwendung findet (vgl. dort **Fig.** 3). Ein solches Modell (dort 320) wird mit bestimmten Wochentagen und Zeiten gespeist, um Wahrscheinlichkeiten einer gewünschten Telefonnummer zu ermitteln. Mit dieser stochastischen Vorbereitung wird eine Spracherkennung kombiniert (dort 330), wofür ein "Integrator" (dort 350) Verwendung findet. Hierbei geht es nur um die Erkennung von Telefonnummern, die ggf. zuvor rückgefragt werden, bevor sie tatsächlich gewählt werden, vgl. dort Spalte **7**, Zeile **11** bis **56**, insbesondere Zeilen **54** bis **56**. Die dort beschriebene Spracherkennung (dort 330) wird als entweder sprecher-unabhängig oder sprecherabhängig beschrieben, wobei eine gewisse Anpassung dieser Spracherkennung an den Sprecher erfolgen kann, wenn er als solcher gespeichert ist (Spalte **7**, Zeile **28** bis **33**).

[0007] Eine alternative Struktur eines Telefonbuchs findet sich in US 5,832,063 (Vysotsky). Es wird dort eine Mischung aus sprecher-abhängigem Erkennungssystem und sprecher-unabhängigem Erkennungssystem vorgeschlagen, wobei ggf. ein Schiedsrichter (dort 406, 254) eine Entscheidung fällen muß, ob eine Telefonnummer gewählt worden ist oder ein Steuersignal (im Sinne eines Command-Wortes) gewünscht war. Es ist die Sprecher-Abhängigkeit bei der Wahl der Telefonnummern vorgesehen, und die Sprecher-Unabhängigkeit bei der Wahl der Kommandos.

[0008] Eine noch weitere Sprachsteuerung findet sich in der WO 95/28790 (Northern Telecom), wo eine Veränderung der HMM dadurch erfolgt, daß sie abhängig von der durch Sprache angesprochenen und freigegebenen Telefonnummer gemacht werden (vgl. dort Anspruch 1 und Seite 6, zweiter Absatz). Schließlich ist aus der US 5,995,929 (Gupta) eine Spracherkennung zur Steuerung einer Telefonanlage bekannt, bei welcher die Wahrscheinlichkeit von Bereichen aufgrund eines Anrufmusters eingestellt werden.

[0009] Aufgabe der Erfindung ist die Schaffung eines besser an einen Benutzer angepaßten Systems zur sprachgeführten Steuerung einer technischen Einrichtung, die beispielsweise eine Telefonanlage sein kann.

[0010] Gelöst wird diese Aufgabe mit Anspruch 1, Anspruch 30, Anspruch 34 oder Anspruch 35.

[0011] Gemäß der Erfindung erfolgt zunächst eine Sprechererkennung. Diese Erkennung läuft darauf hinaus, einen individuellen Sprecher zu erkennen, der mit einem bestimmten Profil in einer Datenbank bereits verfügbar ist. Die Sprechererkennung kann aber auch eine Gruppe (einen Kreis von Personen) identifizieren, die ebenfalls mit einem – allgemeineren, aber schon individualisierten – Profil in der Datenbank verfügbar sind. Diese objektivierte Gruppe umschreibt eine Vielzahl von Sprechern, die aufgrund von objektiven Kriterien dieser Gruppe zugeordnet werden können. Beispiele sind bestimmte Dialekte oder Landessprachen. Weitere Beispiele sind bestimmte psychometrische Eigenschaften. Ebenfalls weitere Möglichkeiten sind Wortschatzeigenschaften im Sinne der Wahl bevorzugter Worte. Schließlich können Satzbaueigenschaften ein objektiviertes Kriterium für eine Sprechergruppe sein. Maßgebend ist dafür aber jeweils das Audiosignal, das entweder dem Individuum oder der objektivierten Gruppe (von Individuen) zugeordnet wird.

[0012] Eine Zuordnung im Sinne einer Authentifizierung muß nicht allein aufgrund eines Abschnitts des Audiosignals erfolgen, es kann auch durch Sekundärindizien zu einer solchen Authentifizierung kommen. Dabei kann aktives Zutun ebenso beteiligt sein, wie Begleitumstände, die eine Authentifizierung erlauben.

[0013] Liegt ein Individuum als spezifischer Autor eines vorliegenden (aktuellen) Audiosignals vor, wird ein diesem zugeordnetes Profil ausgewählt. Gleiches gilt für den Fall, daß eine objektivierte Gruppe von Sprechern festgelegt wer-

den konnte, der dann ein anderes Profil zugeordnet werden kann, welches ausgewählt wird. Es versteht sich, daß das Profil, das einer objektivierten Gruppe zugeordnet werden kann, allgemeiner ist als ein Profil, das einem spezifischen Individuum als Sprecher zugeordnet werden kann.

[0014] Nach Auswahl des Profils wird dieses Profil, das mehrere Parameter einer Spracherkennung im Sinne einer Bedeutungserkennung umfaßt, in eine Erkennungsumgebung geladen, die dazu dient, den ersten Bedeutungsgehalt des Audiosignals zu ermitteln. Mit diesem Laden oder Einbinden, welche Begriffe sinngemäß dieselbe Bedeutung haben, wird die Erkennungsumgebung angepaßt oder vorbereitet. Man kann auch von einer Konfiguration durch Parameter sprechen, die von dem Profil in der Erkennungsumgebung vorgegeben werden. Mit einer solchen Vorgabe (in Sinne einer Vorbereitung oder Anpassung) ist anschließend ein Betrieb der sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technischen Gerätes möglich, das aufgrund des ermittelten Bedeutungsinhaltes des Sprachsignals angesteuert wird. Als Folge der Ansteuerung führt das technische Gerät eine Aktion durch, die in beispielsweise der Bereitstellung oder Wählen einer Telefonnummer, dem Auslösen einer akustischen Rückfrage im Sinne eines Dialoges oder der Schaltung einer Konferenz mit mehreren Teilnehmern besteht. Solche Funktionen sind abhängig von dem Typ des "angesprochenen" Gerätes, das von einer Telefonanlage bis hin zu anderen steuerbaren Geräten vom Wesen her jedes technische Gerät sein kann.

[0015] Die Steuerung des Gerätes erfolgt durch die Erkennung des Bedeutungsgehaltes des Audiosignals, das eine oder mehrere Funktionen des Gerätes gleichzeitig oder nacheinander in vorgegebener Reihenfolge auslösen kann. Eine Aufteilung des Sprachsignals in einen zeitlichen Abschnitt für die Erkennung von Einträgen und in einen zweiten Abschnitt für die Erkennung von Schlüsselworten ist nicht erforderlich.

[0016] Die beschriebene Gruppe von Personen kann aufgrund objektivierter Kriterien festgelegt sein. Diese objektiven Kriterien sind erfaßbare Eigenschaften. Auch lokale "Eigenschaften" im Sinne einer örtlichen Befindlichkeit können verwendet werden, beispielsweise wird eine Gruppe von Personen definiert, die in einem bestimmten Stockwerk eines Gebäudes oder in einem bestimmten Raum eines Hauses sitzt, um von dort zu telefonieren, was schon aufgrund von Nebenstellenanschlüssen als Sekundärindizien zur Authentifizierung führen kann. Hörbare Eigenschaften sind Spracheigenschaften im Sinne von beispielsweise Dialekt oder Fremdsprache.

[0017] Findet das Verfahren kein zugeordnetes Profil für einen individuellen Sprecher oder für eine objektivierte Gruppe, wird von einem Standardprofil ausgegangen (Anspruch 2).

[0018] Dieses Standardprofil enthält in sehr allgemeiner Weise spezifische Kriterien für die Einstellung der Erkennungsumgebung, die weiter unten beschrieben werden. Das Standardprofil kann während des zeitlichen Ablaufs des Analogsignals verändert werden, um zu einem individuellen Profil zu werden. Nach Veränderung wird dieses Profil als ein verfügbares Profil in einer Datenbank abgelegt (Anspruch 3).

[0019] Die Veränderung des Profils kann die Veränderung des mit dem Profil verfügbar gemachten Wortschatzes betreffen (Anspruch 21). Alternativ oder kumulativ kann auch eine Dialogausgabe in ihrer Eigenschaft verändert werden, um sich auf den durch das Profil repräsentierten Benutzer einzustellen und die Dialogausgabe in optischer oder akustischer Form auf das Wissen oder die Fähigkeit des Benutzers einzustellen. Eine weitere Möglichkeit der Änderung des Standard-Profils zum individuellen Profil ist die Vorgabe von Parametern zur Beeinflussung der Wortfolgeerkennung in der Erkennungsumgebung. Hier kann die Anzahl der zur Erkennung zugelassenen Wortfolgen verändert werden, wobei damit sowohl eine Reduzierung, wie auch eine Verlagerung, wie auch eine Erweiterung gemeint ist.

[0020] Eine Rückkopplung durch eine akustische oder optische Signalisierung im Sinne einer Dialogaufforderung kann auch dazu verwendet werden, die Erkennungsumgebung umzustellen, orientiert an den erwarteten Eingaben eines Benutzers (Anspruch 23). Hier kann die Wortfolgeerkennung eingeschränkt werden, um nur noch ganz spezifische Wortfolgen zur Wortfolgeerkennung zuzulassen, insbesondere kann eine Art wiederholte Authentifizierung erfolgen, wenn ein erster Authentifizierungsversuch gescheitert ist und dazu führte, daß ein Standard-Profil in die Erkennungsumgebung eingebunden wurde (Anspruch 23). Schließlich kann mit einer zum Dialog auffordernden Signalisierung ein Wechsel des derzeit in der Erkennungsumgebung eingebundenen (oder geladenen) Profils vorbereitet oder veranlaßt werden (Anspruch 22), wobei anzumerken ist, daß die eigentliche Eingabe des Benutzers hier nicht von der Erfindung umfaßt ist, sondern allein die Signalisierung und die Vorbereitung zur Entgegennahme einer Eingabe zur Erreichung des technischen Effekts der Anpassung an einen Benutzer genügt.

[0021] Aufgrund von Sekundärindizien kann eine Vorauswahl von auszuwählenden Profilen getroffen werden (Anspruch 4). Diese reduzierte Anzahl von verfügbar gemachten Profilen senkt die Wahrscheinlichkeit, daß falsche Profile ausgewählt werden, und erhöht die Geschwindigkeit der Auswahl eines Profils durch die Authentifizierung. Neben statistischen Möglichkeiten können auch die erwähnten Sekundärinformationen zu einer zunächst durchzuführenden Reduzierung des Umfangs der Profile führen, die überhaupt in die engere Auswahl gezogen werden. Ein Beispiel liegt darin, daß eine Nebenstelle nicht zwingend nur von einem Benutzer verwendet wird, sondern mehrere Benutzer in Frage kommen, die diese Nebenstelle verwenden.

[0022] Abhängig von dem zugeordneten und in die Erkennungsumgebung geladenen Profil kann die Steuerungstiefe der Steuerung angepaßt werden. Damit ist gemeint, daß die Reichweite der Steuerungsmöglichkeiten verändert werden kann, abhängig von dem Individuum oder der objektivierten Gruppe. Bestimmte Telefonbucheinträge können bei Anwendung einer Telefonanlage zusätzlich verfügbar gemacht werden, wenn in der Hierarchie höher stehenden Personen authentifiziert werden. Bei Erkennung von unerfahrenen Personen, die auch als ein Profil in der Datenbank im Sinne eines Kreises von Personen verfügbar sind, können nur ganz begrenzte technische Möglichkeiten zur Steuerung des technischen Gerätes verfügbar sein, die beispielsweise wenig komplex sind. Zur Reduzierung der Reichweite der Steuerungsmöglichkeiten (im Sinne der Steuerungstiefe) werden weniger Schlüsselworte aktiv geschaltet, wenn sie in der Bedeutungserkennung der Erkennungsumgebung erkannt werden und dem technischen Gerät zur Ausführung übergeben werden.

[0023] Abhängig von der Authentifizierung kann weiterhin auch ein Bedeutungswandel erfolgen (Anspruch 9, Anspruch 12). Von einer bestimmten Nebenstelle – bei Anwendung auf eine Telefonanlage – kann der Bedeutungswandel den Begriff "mein Chef" erfassen. Abhängig von der Authentifizierung und dem gewählten Profil, das in die Erkennungsumgebung geladen ist, bekommt das Wort "mein" eine unterschiedliche Bedeutung. Entsprechend der durch die "semantische Analyse" gewandelten Bedeutung wird das technische Gerät gesteuert.

[0024] Die semantische Analyse wird durch zumindest einen weiteren Parameter des eingebundenen Profils beeinflußt.

[0025] Die Wortfolgeerkennung, die der semantischen Analyse vorgelagert ist, ist als lexikalische Analyse abhängig vom durch das Profil bestimmten Wortschatz und Satzbau (Syntax). Der Wortschatz wird von einer jeweiligen Stelle im Satz (als Folge von mehreren Worten) abhängig, d. h. an bestimmten Stellen in einem Satz werden nur bestimmte Worte zugelassen und damit erkannt (Anspruch 35, Anspruch 28). Nachdem die für den Erkennungsprozeß zugelassenen Wortfolgen, ggf. auch die zum Dialog auffordernden Signalisierungen, durch das Profil veränderbar sind, paßt sich die Erkennungsumgebung an eine beliebige Vielzahl von Benutzern an, sei es durch Auswählen, Zuordnen und Einbinden eines einem schon bekannten Benutzer zugeordneten individualisierten Profils, dessen weitere Individualisierung und erneute Abspeicherung, oder sei durch Definition von neuen individuellen Profilen, deren Ausgangspunkt sowohl ein Standard-Profil wie auch solche Profile sein können, die auf die objektivierten Gruppen hin angepaßt sind, welche auch als eine Art individualisierte Profile anzusehen sind, die aber nicht so stark individualisiert sind, wie die einzelnen Sprechern zugeordneten Profile. Alle Profile sind in einer Datenbank verfügbar, wobei eine zweite Datenbank vorgesehen sein kann, die der Erkennung und Zuordnung (im Rahmen der Authentifizierung) zugewiesen ist, um die in der ersten Datenbank verfügbaren Einträge den authentifizierten Benutzern zuzuordnen.

[0026] Wird eine neue Benutzergruppe angelegt oder ein neues individualisiertes Profil für einen Benutzer in der Erkennungsumgebung erstellt und anschließend in der ersten Datenbank abgespeichert, erfolgt auch eine Eintragsänderung oder -ergänzung in der zweiten Datenbank, zur Ermöglichung der Zuordnung für spätere akustische Signale, die von dem neuen Benutzer stammen.

[0027] Die Wortfolgeerkennung (Anspruch 10) besitzt auch einen ihr zugänglichen Wortschatz, der durch das Profil vorgegeben wird. Die Wortfolgeerkennung kann mit einer Syntaxerkennung ergänzt sein (Anspruch 11).

[0028] Zur Verbesserung des akustischen Modells kann eine Anpassung dieses Modells abhängig von der Art der Übertragung des Sprachsignals oder von der Art der Aufnahme des Sprachsignals erfolgen (Anspruch 27).

[0029] Von den beschriebenen Systemen, mit den das Verfahren ausgeführt wird, können zumindest zwei, bevorzugt mehrere Pfade ausgebildet werden, die gemäß den Merkmalen (b) und (c) des Anspruchs 1 gestaltet sind, vgl. Anspruch 30. Jeder dieser Pfade ist als eine Funktionslinie beschrieben, die eine jeweils eigene Erkennungsumgebung besitzt. Jeder der Erkennungsumgebungen wird das Sprachsignal identisch zugeführt. Vorgelagert ist nur eine Authentifizierung für alle Erkennungsumgebungen, der dasselbe Sprachsignal zugeführt wird. Abhängig von ihrem hier mehrdimensionalen Ausgangssignal werden den mehreren Erkennungsumgebungen unterschiedliche Profile vorgegeben, die alle dem Sprachsignal zuordnungsfähig erscheinen, nachdem keine eindeutige Festlegung erfolgen kann. Jede Linie arbeitet gesondert, besitzt eine gesonderte Anpassung durch Auswahl und Zuordnung eines eigenen Profils für die jeweils gesonderte Erkennungsumgebung.

[0030] Eine Entscheidungseinrichtung erhält die Ergebnisse der Erkennungsumgebungen, um auszuwählen, welches Erkennungsergebnis dem technischen Gerät zu Steuerungszwecken zugeführt wird. Beispielsweise kann aufgrund einer Schwellenentscheidung eine Bedeutung bevorzugt werden, z. B. durch eine Strahlsuche (eine Suche, die die Bedeutungsinhalte der Erkennungslinien unterdrückt, deren Bewertung unter einer Schwelle liegt). Eine alternative Vorgehensweise ist die Bereitstellung einer ungeraden Zahl von Erkennungslinien (jeweils gesonderte Profilauswahl und Erkennungsumgebung), um die Mehrheit entscheiden zu lassen, welche Bedeutung dem technischen Gerät zur Ausführung weitergegeben wird (Anspruch 32).

[0031] Ein Profil kann Parameter für eine semantische Analyse und/oder eine lexikalische Analyse umfassen. Es können Parameter für ein akustisches Modell noch hinzutreten. Die lexikalische Analyse besteht aus Wortschatz und Satzbau (Syntax, Aufbau des Satzes als Teil der Grammatik). Die semantische Analyse betrifft die Bedeutung von Worten (als Zeichenfolgen), auch die Bedeutung von Wortfolgen, bis hin zum Inhalt eines ganzen Satzes. Bevorzugt ist die Reihenfolge diejenige, ein akustisches Modell vorzulagern, eine lexikalische Analyse folgen zu lassen und eine semantische Analyse hinzuzunehmen (Anspruch 12). Jede dieser drei Funktionsblöcke innerhalb einer Erkennungsumgebung wird von Parametern eines Profils beeinflußt. Unter einer lexikalischen Analyse ist auch ein Wortschatz zu verstehen, der durch das ausgewählte und in die Erkennungsumgebung geladene Profil bestimmt wird. Er kann aus einem vorgegebenen Anteil und einem auf den Benutzer oder die Benutzergruppe zugeschnittenen Anteil bestehen.

[0032] Bei der semantischen Analyse (im Sinne der Adaption der Grammatik) kann der Bedeutungswandel erfolgen (Anspruch 9). Die Abhängigkeit ergibt sich über die Authentifizierung in das ausgewählte Profil, das hinsichtlich der die Semantik betreffenden Parameter auf die Erkennungsumgebung Einfluß nimmt.

[0033] Eine Dialogsteuerung kann zusätzlich vorgesehen sein. Diese Dialogsteuerung arbeitet über eine optische oder akustische Rückkopplung und/oder über eine Rückkopplung innerhalb der Erkennungsumgebung. Eine Rückkopplung über ein Signal, hin zum insoweit tatsächlich erreichbaren Sprecher, führt zu einer Aufforderung des Systems dann, wenn die Bedeutung des vorliegenden akustischen Signals nicht eindeutig erfaßt werden kann oder zuvor ein individuelles Profil nicht zugeordnet werden kann.

[0034] Die optische (am Display) oder akustische Rückkopplung der Dialogsteuerung ist primär orientiert an der Geübtheit des Benutzers. Ist das Profil, das ausgewählt und geladen worden ist, kennzeichnend für einen erfahrenen Benutzer, sind die Rückkopplungen über die akustisch oder optisch sich äußernden Signale schlicht und kurz. Bei geladenem Profil eines ungeübten Benutzers oder bei mehrfachen Fehlern wird die Dialogsteuerung so beeinflußt, daß die Informationsfülle der Rückkopplung vergrößert oder intensiviert wird.

[0035] Im Profil können hinsichtlich der Dialogsteuerung bestimmte Typen vorgegeben sein, die im Sinne von Begriffspaaren (Kooperativität, unkooperativ oder technischer Laie, Systemverständnis) definiert sein können (Anspruch 16, 21). Zumindest ein Parameter des Profils beeinflußt vorteilhaft die Eigenschaft der Dialogsteuerung. Die Ausgabe der Dialogsteuerung erfolgt optisch oder akustisch.

[0036] Ausführungsbeispiele erläutern und ergänzen die Erfindung.

[0037] **Fig.** 1 ist ein schematisches Funktionsbild einer Steuerungseinrichtung zur Steuerung des technischen Gerätes **52**, das im Beispiel als Telefonanlage ausgestaltet sein kann.

[0038] **Fig.** 2 ist eine Steuerungseinrichtung **1**a, bei der jeweils eine Profilauswahl **31**A und eine Erkennungsumgebung **4**A einen Erkennungspfad bildet, von denen mehrere parallel geschaltet sind.

[0039] **Fig.** 3 veranschaulicht den inneren Aufbau einer Erkennungsumgebung **4**, angesteuert von einem Sprachsi-

gnal s_a und beeinflußt (vorbereitet, konfiguriert oder angepaßt) von einer Profilauswahl **31**.

[0040] Beschrieben werden soll das Verfahren zum Vorbereiten, Betreiben und Anpassen einer sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technischen Gerätes anhand einer Telefonanlage, die als Gerät **52** vorstellbar ist, das gemäß **Fig.** 1 Befehle über die Steuerungsleitung **20** erhält und in technische Funktionen umsetzt, die als "Action" bezeichnet sind. Solche technischen Funktionen können das Wählen einer Telefonnummer, das Einleiten einer Konferenzschaltung, das Sperren eines Zugangs, das Umschalten auf ein anderes Telefon oder sonstige Funktionen sein, die von heutigen Telefonanlagen in einer Großzahl angeboten werden, allerdings jeweils wenig komfortabel gesteuert über Knopfdruck-Sequenzen oder Sondertasten.

[0041] Die Sprachsteuerung bietet eine große Flexibilität, eine leichte Bedienbarkeit und eine hohe Funktionalität. Ein Leistungskriterium bei der Spracherkennung ist in erster Linie die Erkennungssicherheit. Sie wird geringer, wenn die Sprecher variieren, also eine Erkennung unabhängig vom Sprecher bereitgestellt werden soll. Daneben ist Erkennungsgeschwindigkeit ein wichtiges Kriterium. Sie hängt von der Komplexität der zugelassenen Wortfolgen ab. Je weniger Wortfolgen berücksichtigt werden müssen, desto schneller die Erkennung, aber vor allem desto höher die Erkennungssicherheit. Ein Einflußkriterium auf die Dauer und die Sicherheit der Erkennung ist der Wortschatz und die grammatikalische Komplexität (= Perplexität) des Sprachsignals. Geht man deshalb von freiem Dialog mit kontinuierlicher Sprache aus und möchte jeden Sprecher zulassen, so muß das System Sicherheit mit Geschwindigkeit paaren, bei gleichzeitig komplexer Steuerungsmöglichkeit des technischen Gerätes. Gegen eine drastische Reduzierung der zugelassenen Wortfolgen spricht aber, daß die Äußerung für eine erfolgreiche Erkennung natürlich auch dabei sein muß, um auch benutzerspezifische Wortfolgen erkennen zu können.

[0042] Als eine Linie, mit der die Erkennungsumgebung der **Fig.** 3 skizzierbar ist, ist die Reihenfolge aus einem digitalen Sprachsignal, einer Berechnung der akustischen Merkmale des Sprachsignals, die Bestimmung der besten Wortfolge durch akustisches und syntaktisches Wissen sowie die semantische Analyse der Wortfolge anzugeben. Die semantische Analyse ergibt die Bedeutungserkennung, also den Inhalt des Sprachsignals, das über die Steuerleitung **20** eine Aktion des Gerätes **52** bewirkt.

[0043] Eine Dialogstruktur durch Ausgabe eines rückkoppelnden Sprachsignals **51** zum Sprecher kann zusätzlich vorgesehen sein. Sie folgt gewissen Strategien, die weiter unten erläutert werden sollen.

[0044] Ausgangspunkt für die Erkennung des Bedeutungsgehaltes eines Sprachsignals ist das Signal selbst. Dieses Signal s_a wird der Steuerungseinrichtung **1** an zwei Stellen zugeführt, einem Eingang **9** der Erkennungsumgebung **4** und an einem Eingang **10** der Authentifizierung **2**. Das Signal wird von einem Eingang oder einer solchen Schaltung **11** bereitgestellt, der entweder ein Mikrophon **11a** oder ein digital oder analog gespeichertes Signal oder ein über ein Netz übertragenes Signal, bei nicht anwesendem Sprecher, vorgelagert ist. Der Authentifizierung **2** wird Sekundärinformation **12** zugeführt. Die Schaltung zur Authentifizierung gibt ein Ausgangssignal "a" ab, mit dem ein Profil ausgewählt wird, was in einer Auswahlschaltung **31** mit Zuordnungssektion **31*** erfolgt, die Zugriff auf eine Datenbank **32** besitzt, in der eine Vielzahl von Profilen P_i abgelegt sind.

[0045] Mit der Profilauswahl **31** wird ein ausgewähltes Profil P_i, wobei $i = 1 \ldots n$, in die Erkennungsumgebung **4** geladen oder eingebunden. Das Einbinden oder Laden ist so zu verstehen, daß bestimmte Parameter der in der Erkennungsumgebung vorhandenen Funktionselemente verändert werden. Das eingebundene Profil ist mit 33 bezeichnet, es ist ein Element der gespeicherten individualisierten Profile P_i. Ein weiteres gespeichertes Profil ist das Standard-Profil P_x, das weiter unten erläutert wird.

[0046] Die Funktion der Authentifizierung und der Auswahlschaltung **31** mit Zuordnungssektion **31*** kann zusammengefaßt werden. Die Zuordnung kann über eine zweite Datenbank geschehen, die auch als Teil der ersten Datenbank **32** angesehen werden kann. Die Authentifizierung kennzeichnet einen Benutzer oder eine objektivierte Gruppe von Benutzern, um zugehörig ein Profil aus der Haupt-Datenbank **32** zu entnehmen, welche Zuordnung über die Hilfs-Datenbank **31*** erfolgt. Das ausgewählte Profil **33** wird in die Erkennungsumgebung **4** geladen.

[0047] Der Erkennungsumgebung wird das Sprachsignal s_a über den genannten Eingang **9** auch zugeführt, so daß es mit den eingestellten Parametern des Profils **33** bearbeitet werden kann. Eine Darstellung des inneren Aufbaus der Erkennungsumgebung ist in **Fig.** 3 gezeigt und wird später erläutert.

[0048] Aus der Erkennungsumgebung folgt ein Signal **20**, welches das technische Gerät **52** steuert.

[0049] Die Erkennungsumgebung arbeitet zusammen mit einer Dialogsteuerung, die eine Signalausgabe **51** einsetzt. Diese Display- oder Sprachausgabe **51** ist nur dann sinnvoll, wenn das Signal angemessenem zeitlichen Rahmen einen Benutzer erreicht. Über die Dialogsteuerung wird eine Signalausgabe erzeugt, die an den Sprecher zurückgekoppelt werden kann, was auch über eine große Entfernung oder über ein Netzwerk und auch zeitlich versetzt möglich ist. Bei Telefonanlagen ist die Erreichbarkeit durch Rückkopplung unmittelbar ersichtlich, das Dialogsignal wird in die Hörerleitung direkt eingeblendet, während das Mikrophon als Quelle für das bearbeitete Signal s_a dient. Gleiches kann auch auf optischem Wege durch ein Display erfolgen.

[0050] Ein Benutzer, der ein akustisches Signal zum Mikrophon **11a** gibt, spricht die Authentifizierung **2** an. Die Authentifizierung bestimmt den Sprecher. Sie versucht zunächst, aufgrund des gesprochenen Wortes eine Zuordnung zu finden, ob ein in der Datenbank **32** verfügbares individuelles Profil zu dem Sprachsignal s_a am Eingang **10** zugeordnet werden kann. Sekundärinformationen können stützend oder alleinig herangezogen werden, beispielsweise als Nebenstelleninformation, die angibt, welcher Benutzer das Sprachsignal abgibt. Der Benutzer kann sich auch aktiv selbst authentifizieren, was über eine Dialogsteuerung mit Sprachausgabe **51** möglich ist. Steht kein spezielles Profil in der Datenbank **32** zur Verfügung, das von der Zuordnung **2**, **31**, **31*** eindeutig zugeordnet werden kann, wird das Standard-Profil P_x verwendet. Kann anhand des Sprachsignals, anhand der Sekundärinformation oder anhand von Eigenidentifizierung eine Zuordnung zu einem Kreis von Personen erfolgen, für den ein gespeichertes Profil zur Verfügung steht, wird dieses Profil als gruppenspezifisches Profil, betreffend eine Gruppe von Personen, ausgewählt. Das ausgewählte Profil **33** wird in die Erkennungsumgebung **4** geladen und beeinflußt hierbei die an **Fig.** 3 näher zu beschreibenden Parameter. Bei der Durchführung der Spracherkennung in der Erkennungsumgebung **4** ist durch das ausgewählte Profil vorgegeben, wie die Erkennung erfolgen soll und welche Parameter dazu in den akustischen Modellen, bei dem Wortschatz anhand von geladenem Wortschatz (Wörterbüchern) und in der semantischen Analyse Anwendung finden.

[0051] Das Standard-Profil P_x kann dann, wenn es in die Erkennungsumgebung **4** geladen ist, modifiziert werden. Das Standard-Profil P_x bleibt dabei in der Datenbank **32** un-

verändert, nur das Abbild wird in der Erkennungsumgebung geändert, um nach der Änderung als ein individualisierteres Profil P_i neu abgespeichert zu werden. Die Veränderung kann dabei alle Bereiche der Parameter erfassen, die an der Profilumgebung näher erläutert werden, so Parameter für die akustischen Modelle, Parameter für die lexikalische Analyse (Wortschatz und Syntax) sowie Parameter für die semantische Analyse (Bedeutungswandel). Weitere Parameter für die Anpassung der Dialogsteuerung können ebenfalls verändert werden. Das Rückspeichern eines erzeugten neuen individuellen Profils in die Datenbank **32** sorgt auch für eine Anpassung der Einträge in der Hilfs-Datenbank **31***, so daß die Authentifizierung **2** in Verbindung mit der Zuordnung **31** eine Auswahl auch zukünftig vornehmen kann, wenn der Benutzer später erneut auftritt.

[0052] Wird in der Erkennungsumgebung ein schon vorhandenes individuelles Profil weiter individualisiert, ergänzt, erweitert oder geändert, so wird dieses Profil an den ursprünglichen Platz in der Datenbank zurückgespeichert und die Hilfs-Datenbank **31*** als Zuordnungssektion nicht erneut aktualisiert.

[0053] Die **Fig.** 3 veranschaulicht den inneren Aufbau der Erkennungsumgebung **4**. Das Sprachsignal s_a wird dem Eingang **9**a zugeführt. Das ausgewählte Profil mit den mehreren Parametern wird an einem Eingang **8**a zugeführt und stellt über 41' Parameter des akustischen Modells **41**, der lexikalischen Analyse **42** und der semantischen Analyse **43**, die in Serie geschaltet sind, ein. Eine Dialogsteuerung **44** ist der semantischen Analyse nachgeordnet und kann alle drei genannten Funktionsblöcke beeinflussen und von Parametern des Profils über 41' beeinflußt werden. Sie steuert auch eine Display- oder Sprachausgabe **51**, mit der eine Rückkopplung zum Urheber des Sprachsignals s_a erfolgt. Aus der Dialogsteuerung **44** oder aus der semantischen Analyse **43** direkt ergibt sich ein Steuersignal **20**, das das elektrische Gerät **52** zu einer Aktion veranlaßt, wenn es über den Ausgang **7**a zum Eingang **52**a des zu steuernden Gerätes **52** übertragen wird. Die Aktion kann auch eine innere Aktion im Rahmen des technischen Gerätes **52** selbst sein.

[0054] Folgend sind einzelne Funktionen der Funktionsblöcke **41**, **42**, **43** und **44** erläutert.

[0055] Das akustische Modell oder die akustischen Modelle **41** werden adaptiert durch Vorgabe von Parametern aus dem Profil, wobei ein Abschnitt des Gesamtprofiles diese Parameter enthält. Möglich ist eine unüberwachte Adaption aller oder eines Teils der akustischen Modelle anhand des maximum a posteriori Kriteriums. Akustische Modelle als solches sind Stand der Technik und können als HMM (Hidden Markov Model) oder als ein Sprachmodell Einsatz finden. Die Auswahl der einzustellenden Parameter geschieht anhand des zur Verfügung stehenden Trainingsmaterials des Sprecher, das zur Bildung des Profils geführt hat. Es ist bereits mit einer Äußerung eine Anpassung der akustischen Modelle möglich, dabei werden nur Mittelwerte der verwendeten Ausgabeverteilungsdichten verändert.

[0056] Ist als Profil das Standardprofil P_x ausgewählt, bei dem noch keine spezifische Zuordnung zum Sprachsignal vorliegt, so kann dieses Standardprofil im Laufe des Arbeitens des akustischen Modells **41** angepaßt werden und zur Abspeicherung eines sich ergebenden neuen Profils führen. Die Parameter werden in Form von Merkmalen gespeichert, um möglichst platzsparend speichern zu können.

[0057] Die Anpassung des akustischen Modells durch Einstellen von Parametern betrifft in der Regel die Einstellung von Merkmalen. Sind genügend Daten vorhanden, können neben den Mittelwerten auch Kovarianzen angepaßt werden.

[0058] Zur Verbesserung der Erkennung im akustischen Modell kann ein Anpassen auch insoweit erfolgen, daß das Sprachmodell an die Art der Übertragung oder die Art der Aufnahme des aktuellen Audiosignals angepaßt wird. Die Art der Übertragung spricht den Kanal an, über den das Audiosignal übertragen wurde, so das Festnetz, Mobilfunk oder Satellitenfunk. Die Art der Aufnahme spricht die Möglichkeit an, über eine Freisprechanlage oder ein Mikrophon zu sprechen, ebenso könnte die Digitalisierung berücksichtigt werden, oder die Verwendung eines analogen Audiosignals. Es ergibt sich in jedem Fall eine bessere akustische Auswertung im akustischen Modell **41**.

[0059] Die Funktionseinheit **42**, die auf die akustischen Modelle **41** folgt, enthält über das ausgewählte und geladene Profil einen vorgegebenen Wortschatz. Ein vorgegebener Wortschatz enthält spezifisch ausgewählte Wörter, die für das technische Gerät eine Bedeutung haben, und andere Worte.

[0060] Im Falle einer Telefonanlage können beispielsweise folgende Wörter als Schlüsselwörter in einem festen System-Wortschatz verankert sein, der nicht an den Sprecher angepaßt wird, sondern allenfalls an eine objektivierte Gruppe von Sprechern, also einen Kreis von Sprechern, die eine spezifische gemeinsame Eigenschaft besitzen, die erfaßbar ist. Ein Beispiel ist eine Wortschatzdefinition des festen System-Wortschatzes für eine Personengruppe, wie alle Bayern oder alle Berliner oder alle Sachsen.

[0061] Als Fundus für den festen Wortschatz (System-Wortschatz) können für den Betrieb einer Telefonanlage die Begriffe "Telefonnummer", "Nebenstelle", "Anschluß", "Nummer", "Umleitung", "umleiten", "Ruf", "Telefonbuch", "Konferenz", "Schaltung", "Verbindung", "verbunden" verwendet werden. Kontextworte wie "die", "das", "den", "dem", "unser" "unseren", "in", "im", "bei", "von" sind ein zweiter Bestandteil des System-Wortschatzes. Ein dritter Bestandteil besteht aus Gruppenzuordnungen, wie "Firma", "intern", "Abteilung". Ein vierter Bestandteil des System-Wortschatzes kann in Tätigkeiten bestehen, die vorgegeben werden, namentlich "werden", "aufheben", "navigieren", "auswählen", "Auswahl". Eine schließlich letzte Anteilsgruppe des System-Wortschatzes kann Schlüsselinformationen enthalten, die der Sprecher wünscht, wie "kein", "keine", "Ahnung", "weiß", "ich", "möchte", "will", "sprechen", "bitte". Es versteht sich, daß auch alle nötigen Ziffern z. B. zwischen 0 und 99 Gegenstand des System-Wortschatzes sind.

[0062] Ein dynamischer Wortschatzanteil wird von dem Profil, das ausgewählt und vorgegeben wird, bestimmt. Im dynamischen Wortschatz können Personen verzeichnet sein, mit denen der Sprecher, der über das Profil definiert ist, nicht sprechen will oder sprechen möchte. In einem ergänzenden Verfahren ist es möglich, nach einer durchgeführten Wahl über eine Telefonnummer eine oder mehrere Personen auch bisher nicht gewählter Anschlüsse mit in das Profil zu übernehmen, wobei als Kriterium eine gemeinsame Eigenschaft Anwendung finden kann.

[0063] In einer einfachen Variante kann auch einfach von einem definierten Wortschatz ausgegangen werden, der von dem Profil der lexikalischen Analyse **42** zugeordnet wird. Die lexikalische Analyse arbeitet mit dem ihr zugeordneten Fundus an Worten. In einer genaueren Aufteilung können die oben beschriebenen System-Worte zumindest teilweise übernommen werden; es kann auch eine Zuordnung bestimmter System-Worte zu einem bestimmten Profil erfolgen, während andere System-Worte zu einem anderen Profil zugeordnet werden. Bereits hierdurch kann die zuvor beschriebene Steuerungstiefe des zu steuernden Gerätes **52** profilabhängig werden.

[0064] Die lexikalische Analyse **42** durch den Wortschatz

wird ergänzt durch eine Adaption der Grammatik im Sinne der Syntax. Äußerungen des Benutzers werden, wie bereits oben erwähnt, zum einen in Form akustischer Merkmale und in Form einer erkannten Kette gespeichert. Ab einer gewissen Anzahl von Äußerungen können Übergangswahrscheinlichkeiten zwischen Wörtern des zuvor beschriebenen Wortschatzes angepaßt werden, das die Produktionswahrscheinlichkeit der beobachteten Benutzeräußerungen maximiert. Dabei können u. U. ganze Pfade in der Syntax vollkommen ausgeschaltet werden.

[0065] Ein solches Ausschalten findet dann statt, wenn – für eine Telephonanlage – nur noch Äußerungen der Form "Herrn Müller der Firma Y" zugelassen werden, aber solche Äußerungen, wie "ich möchte Herrn Müller der Firma Y sprechen" nicht mehr zugelassen werden. Aus Gründen der Robustheit sollte jedoch die Möglichkeit gegeben werden, alternative Sprechformen zu berücksichtigen, allerdings schlechter zu bewerten.

[0066] Die für die Syntaxerkennung in 42 zugelassenen Wortfolgen können durch das Profil verändert sein. Die für die Erkennung zugelassenen Wortfolgen können auch durch den Systemzustand, insbesondere den Eingriff der Dialogsteuerung **44** verändert werden, wenn nur eine begrenzte Anzahl von Bedeutungen zugelassen oder erwartet werden. Der Wortschatz kann auch von der jeweiligen Stelle in der Folge mehrerer Worte abhängig sein, d. h. an bestimmten Stellen in einer Wortfolge werden nur bestimmte Worte aus dem durch das Profil zugeordneten Wortschatz zugelassen und damit erkannt. Die Einschränkung kann also sowohl die für das nächste Wort zur Verfügung stehenden Worte aus dem Wortschatz betreffen, wie auch die Menge der zur Erkennung zugelassenen Wortfolgen beeinflussen.

[0067] Eine semantische Analyse **43** ist der lexikalischen Analyse nachgeschaltet. Auch sie wird beeinflußt durch einen Ausschnitt der Parameter aus dem ausgewählten und geladenen Profil. Bei der semantischen Analyse werden Bedeutungen ermittelt, die einem Wort, einer Wortfolge oder einem ganzen Satz entspringen. Es können Wandel in der Bedeutung einfließen. Die Authentifizierung hat Einfluß auf die Profilauswahl und die diesbezüglich relevanten Parameter beeinflussen die semantische Analyse. Ein bestimmter Begriff einer spezifischen Person oder einer Personengruppe wird in der semantischen Analyse neu belegt, so der Begriff "mein" abhängig von dem Anrufer, oder der Begriff "Empfänger" für eine spezifisch wechselnde Person. Auch der Begriff "mein Mann" kann sich in der Bedeutung wandeln.

[0068] Eine Dialogsteuerung **44** arbeitet der semantischen Analyse nachgeordnet und beeinflußt alle drei Funktionen **41**, **42**, **43**. Die semantische Analyse kann dann, wenn sie ein eindeutiges Ergebnis des Sinngehalts des Sprachsignals ermittelt, auch direkt den Ausgang **7**a speisen.

[0069] Die Dialogsteuerung ist vorgelagert, um Fehler zu vermeiden, Redundanz zu erlauben und eine Anpassung der Erkennungsumgebung **4** an den Sprecher, respektive das ihm zuzuordnende Sprachsignal s_a am Eingang **9**a zu erlauben. Treten bei der Erkennung mehrmals Fehler auf, so wird in einen eingeschränkten Dialog umgeschaltet. Beispielsweise geschieht das dadurch, daß nur noch nach einem einzigen Eintrag im Rahmen einer akustischen Rückkopplung über die Sprachausgabe **51** und Lautsprecher **51**a gefragt wird. Es wird dann eine Antwort von dem Sprecher über das Sprachsignal s_a am Eingang **9**a erwartet. Auf diese Antwort kann die Dialogsteuerung die akustischen Modelle, die lexikalische Analyse und die semantische Analyse voreinstellen.

[0070] Die zuvor beschriebene Umschaltung auf einen eingeschränkten Dialog kann auch mehrstufig erfolgen, also in der höchsten Stufe kann dem Sprachsignal ein großer Freiraum gegeben werden, in einer niedrigeren Stufe nur Vorname, Nachname und Firma, schließlich in einer letzten Stufe nur ein einzelner Eintrag. Diese Stufenauswahl kann über einen gesamten Dialog hinweg gleich bleiben.

[0071] Eine weitere Möglichkeit der Steuerung liegt in der Länge oder in dem Detail bzw. der Detaillierung der Sprachausgabe **51**, **51**a. Ein vertrauter Benutzer braucht wenig Informationen über die Sprachausgabe durch die Dialogsteuerung, um das System zielgerichtet anzusprechen und zu beeinflussen. Für einen Anfänger werden längere Informationen als Prompts vorgesehen, um ihm genauer vorzugeben, was von ihm als Sprachsignal am Eingang **9**a gewünscht wird. Die Abhängigkeit der Dialoglänge hängt von der Vertrautheit des Benutzers ab, was ein Merkmal in dem Profil sein kann, das ausgewählt und über 31, 41' am Eingang **8**a eingestellt wird.

[0072] Aus **Fig.** 2 ist eine Parallelverarbeitung **1**a ersichtlich. Sie arbeitet mit einem System gemäß **Fig.** 1, dessen Erkennungsumgebung **4** so gestaltet ist, wie **Fig.** 3 zeigt. Es sind mehrere parallele Linien F_A F_B, ... F_J, F_X vorgesehen, die beliebig erweiterbar sind, wobei J = A, B, ... X. Jedes Ausgangssignal A, B, C, ... X einer Erkennungsumgebung **4**A, **4**B, **4**C, ... **4**X einer der Linien F_J wird einer Entscheidungsvorrichtung **6** zugeführt, die ein Ausgangssignal über einen Leitung **21** abgibt, das demjenigen der Leitung **20** aus **Fig.** 1 entspricht, zur Steuerung des Geräts **52** über einen Eingang **52**a. Jede Erkennungsumgebung wird von einer Profilzuordnung **31**A, **31**B ... **31**X angesteuert, die aus der Datenbank **32** ein jeweiliges Profil abhängig von der Authentifizierung **2** ausliest und der zugehörigen Erkennungsumgebung über 8a, 8b, ... 8x zuführt. Das Sprachsignal s_a wird der Authentifizierung und allen Erkennungsumgebungen gleichermaßen über einen jeweiligen Eingang **9**a, **9**b ... zugeführt. Das Authentifizierungssignal a* wird allen Auswahlschaltungen **31**A, **31**B ... gleichermaßen zugeführt. Die Authentifizierung **2** selbst wird so über den Eingang **10** angesteuert, wie in **Fig.** 1 gezeigt. Die Authentifizierung sorgt dafür, daß die zugeordneten Profile nicht dieselben sind.

[0073] Alle Zuordnungsschaltungen **31**A, **31**B, ... greifen auf die Datenbank **32** zu; für die Zuordnung kann eine Hilfs-Datenbank entsprechend der Sektion **31*** von **Fig.** 3 herangezogen werden, die entweder der Zuordnungsschaltung oder der Authentifizierung **2** funktionell zugeordnet ist. Die mehrdimensionale Ausgestaltung des Authentifizierungssignals a* sorgt dafür, daß jede Bedeutungserkennung **4**A, **4**B, ... ein eigenständiges Profil erhält, so daß alle an der Parallelverarbeitung beteiligten Erkennungslinien mit unterschiedlichen Profilen arbeiten.

[0074] Die gleichzeitige Analyse über verschiedene eingestellte Profile in den verschiedenen Erkennungsumgebungen, die alle parallel arbeiten, erlaubt eine genauere Anpassung und eine bessere Erkennung eines schwierigen Sprachsignals s_a. Während des Sprachsignals arbeiten alle Erkennungen mit ihren akustischen Modellen parallel, nur unterschiedlich dadurch, daß andere Merkmale (Mittelwerte oder Kovarianzen) eingestellt sind. Das Ausgangssignal A, B oder X einer Bedeutung wird über die entsprechende Leitung in der Entscheidungsvorrichtung **6** auf die Mehrheit oder auf einen Schwellenwert hin überprüft. Das Ausgangssignal auf Leitung **21** steuert das technische Gerät **52** über seinen Eingang **52**a. Die Schwellenüberprüfung kann durch eine Strahlsuche erfolgen (eine Suche, die alle Erkennungspfade unterdrückt, deren Bewertung unter der Schwelle liegt, wobei die Schwelle durch die Bewertung des zu dieser Zeit besten Pfades liegt). Es ist wahrscheinlich, daß nach einer kurzen Zeit nur noch solche Pfade innerhalb eines Systems nach **Fig.** 2 wirksam sind. Dadurch werden im weite-

ren Verlauf der Spracherkennung alle anderen Profile, deren Ausgangssignal unter der Schwelle liegt, nicht mehr ausgewertet.

[0075] Der Entscheidungseinrichtung **6** kann auch eine Signalausgabe **51**, **51**a, die akustischer Natur oder als Display optischer Natur ist, ansteuern. Mit ihr können Rückkopplungen angestoßen werden, wenn sich bei der Festlegung der für die Steuerung maßgebenden Auswahleingänge A, B, etc. Differenzen ergeben, die es notwendig erscheinen lassen, eine Rückfrage einzuleiten, eine Information zur Bedienung auszugeben oder lediglich Statussignale verfügbar zu machen.

[0076] Ein Profil ist bislang als in der Datenbank **32** vorhanden angesehen worden. Die Erstellung eines solchen Profils war oben bereits skizziert. Sie erfolgt während des Auswertens eines akustischen Signales s_a im akustischen Modell, in der lexikalischen Analyse und in der semantischen Analyse. Das eingangs angesprochene Standardprofil kann so beeinflußt werden und rückgespeichert werden, zur Bildung eines neuen Profils. Diese Rückspeicherung kann sowohl bei **Fig.** 1 erfolgen, wenn ein einzelner Erkennungspfad **31**, **4** Anwendung findet, wie auch bei **Fig.** 2, wenn drei Erkennungspfade F_A, F_B und F_X Anwendung finden.

[0077] Sind die mehreren Erkennungspfade gemäß **Fig.** 2 vorgesehen, kann auch eine alternative Bildung eines neuen Profils gewählt werden. Bleiben bei dieser Art der Verarbeitung des Sprachsignals am Ende noch mehrere Pfade mit unterschiedlichen Profilen aus der Datenbank übrig, so werden die statistischen Eigenschaften entsprechend der statistischen Bewertung der besten Pfade innerhalb der verschiedenen Interessensprofile aufsummiert, und es entsteht ein neues Profil, das abgespeichert wird.

[0078] Die Bildung eines neuen Profils kann auch die Änderung eines bestehenden Profils sein, das an seinen Speicherplatz zugespeichert wird, unter Berücksichtigung der sich während der Auswertung der Sprache ergebenden Änderungen in den Merkmalen gemäß den Einzelfunktionen **41**, **42**, **43**, ggf. auch unter Berücksichtigung von Änderungen in der Einrichtung der Dialogsteuerung **44**.

[0079] Eine Vorgehensweise bei der Bildung eines neuen Profils kann bei der Anwendung der Struktur nach **Fig.** 2 darin liegen, eine Erkennungslinie, beispielsweise F_X, immer mit dem Standardprofil zu betreiben, während das gleiche Sprachsignal über ein anderes ausgewähltes Profil aus der Datenbank **32** in zumindest einem der anderen Pfade nach Maßgabe der Authentifizierung **2** und der Zuordnung **31** ausgewertet wird. Das Ergebnis des Standardprofils kann dann verglichen werden mit dem Ergebnis des eigentlich ein besseres Ergebnis versprechenden Spezialprofils.

[0080] Wird in einer Parallelverarbeitung regelmäßig ein Standardprofil P_x verwendet, braucht diese Profilauswahl nicht von der Authentifizierung **2** mit angesteuert zu werden, sondern bleibt unabhängig davon.

[0081] Neben einer Anpassung des Standardprofiles an einen Sprecher kann auch eine Verbesserung des Standards selbst erfolgen, durch Ergänzung von Merkmalen der akustischen Modelle, der lexikalischen Analyse oder der semantischen Analyse. Dieses neu gebildete Standardprofil wird in der Datenbank **32** so zurückgespeichert, daß es für alle parallelen Profilauswahlen als Standardprofil wieder verfügbar ist.

[0082] Erfolgt die parallele Verarbeitung regelmäßig mit einem Standardprofil kann auch die durch Sekundärinformation fehlgeleitete Vorauswahl eines Profils behoben werden. Ruft über eine Nebenstelle regelmäßig derselbe Sprecher an, und wird diese Nebenstelle einmal von einem anderen Benutzer verwendet, so wäre die Auswahl über die Sekundärinformation im Rahmen der Authentifizierung **2** unrichtig. Hier hilft die standardmäßig verwendete Erkennung über das Standardprofil im Erkennungspfad F_X, deren Erkennungsergebnis X im automatischen Entscheider **6** über einen Schwellenwert oder über eine Gütefunktion mit dem Erkennungsergebnis des eigentlich besseren Profils verglichen wird. Die Entscheidungseinrichtung legt fest, welche Erkennung sicherer funktioniert hat, und nimmt das entsprechende Profil aus dem Speicher **32**.

Patentansprüche

1. Verfahren zum Vorbereiten, Betreiben oder Anpassen einer sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technisches Gerätes (**52**), wie eine Telefonvermittlung oder eine Telefonanlage; wobei ein Audiosignal (s_a) aus zumindest einem von einem Sprecher abgegebenen Wort, insbesondere mehreren aufeinanderfolgenden Worten, welche einen Bedeutungsgehalt besitzen und insoweit erkannt werden sollen, einem ersten Signaleingang (**10**) einer Authentifizierungseinrichtung (**2**) zugeführt werden; wobei
   (a) eine Sprechererkennung aufgrund eines Authentifizierungsversuchs (**2**) erfolgt, insbesondere durch einen zeitlichen Abschnitt des Audiosignals (s_a), durch Erkennen einer Selbstauthentifizierung mittels aktiven Zutuns, oder durch Auswerten von Sekundärindizien (**12**), wie eine bekannte Telefon-Nummer, ein Nebenstellen-Anschluß, um einen Sprecher als Individuum oder eine objektivierte Gruppe von Sprechern festzulegen, der der Sprecher durch objektivierte Kriterien des auf ihn zurückzuführenden Audiosignals zuzuordnen ist, und ein entsprechendes Ausgangssignal (a, a*) abzugeben;
   (b) ein zum festgelegten Sprecher oder der objektivierten Gruppe korrespondierendes Profil (**33**) aus einer Vielzahl gespeicherter Profile (32, P_i) ausgewählt (**2**, **31**) wird, gestützt auf das Ausgangssignal (a, a*) der Authentifizierung;
   (c) das ausgewählte Profil (**33**) in eine Erkennungsumgebung (**4**) eingebunden oder geladen wird, um die Erkennungsumgebung auf den festgelegten Sprecher bzw. die objektivierte Gruppe hin anzupassen; wobei
   (d) jedes der gespeicherten Profile (P_i) und das eingebundene oder geladene Profil (**33**) Parameter enthält, zur Beeinflussung zumindest einer in der Erkennungsumgebung (**4**) vorgesehenen Wortfolgeerkennung (**42**).

2. Verfahren nach Anspruch 1, bei dem nach dem Authentifizierungsversuch, der kein Profil für das aktuelle Sprachsignal als korrespondierend ergibt, ein Standardprofil (P_x) ausgewählt wird und in die Erkennungsumgebung eingebunden wird (**33**).

3. Verfahren nach Anspruch 2, bei dem das in der Erkennungsumgebung geladene Standard-Profil (P_x) während einer zeitlichen Dauer des Sprachsignals (im folgenden "aktuelles Audiosignal") auf dieses Sprachsignal hin angepaßt wird, um zu einem individuellen Profil (P_i) zu werden, das abgespeichert wird (**32**).

4. Verfahren nach Anspruch 1, bei dem vor dem Auswählen oder Zuordnen eines gespeicherten Profils eine reduzierte Anzahl wahrscheinlicher Profile, die sich im Umfang der gespeicherten Profile befinden, als Auswahlmöglichkeiten bereitgestellt werden, um eine Vorab-Auswahl von – vermutlich zuzuordnenden – Profilen als einen eingeschränkten Auswahlbereich zur Verfügung zu stellen, und daraus ein an das Audiosi-

gnal am besten angepasste Profil auszuwählen und in die Erkennungsumgebung (4) zu laden (33).

5. Verfahren nach Anspruch 1, bei dem das ausgewählte und eingebundene (33) Profil zumindest ein HMM oder ein Sprachmodell für ein akustisches Modell (41) als ersten Erkennungsabschnitt in der Erkennungsumgebung (4) umfaßt.

6. Verfahren nach Anspruch 1, bei dem abhängig von dem ausgewählten und geladenen (33) Profil mehr oder weniger Schlüsselworte oder mehr oder weniger Informationseinträge (gespeicherte Informationen) zugelassen werden oder über das Audiosignal gesteuert auswählbar sind, um die Steuerungstiefe der Steuerung des technisches Gerätes (52) von dem aktuellen Audiosignal und damit von dem authentifizierten Sprecher abhängig zu machen.

7. Verfahren nach Anspruch 1, wobei die objektivierte Gruppe ein Kreis von Personen ist, der zumindest eine gemeinsame, erfaßbare Eigenschaft besitzt, die eine objektivierte Zuordnung erlaubt, ob der über das Audiosignal zugeordnete Sprecher zu dem Kreis von Personen zuzuordnen ist oder nicht.

8. Verfahren nach Anspruch 7, wobei die Eigenschaft hörbar, meßbar oder durch einen lokalen Zustand der Befindlichkeit definiert ist.

9. Verfahren nach Anspruch 1, wobei abhängig vom Ausgangssignal (a, a*) der Authentifizierung (2) der Bedeutungsgehalt eines von der Erkennungsumgebung erkannten objektiven Wortes oder Wortfolge im Audiosignal (s_a) einem Bedeutungswandel unterworfen wird, um das technische Gerät (52) entsprechend der gewandelten Bedeutung zu steuern.

10. Verfahren nach Anspruch 1, wobei die Wortfolgeerkennung (42) in der Erkennungsumgebung über einen durch das eingebundene Profil (33) vorgegebenen Wortschatz verfügt.

11. Verfahren nach Anspruch 1 oder 10, wobei die Wortfolgeerkennung eine Syntaxerkennung beinhaltet.

12. Verfahren nach Anspruch 1 oder 10, wobei der Wortfolgeerkennung (42) in der Erkennungsumgebung (4) eine semantische Analyse (43) nachgeordnet ist, die von zumindest einem der Parameter des eingebundenen Profils (33) beeinflußt wird.

13. Verfahren nach Anspruch 1 oder 10, wobei ein Profil (P_i) Parameter enthält, und zwar für die Vorgabe eines Wortschatzes aus einem vorgegebenen Anteil von Worten eines Systemwortschatzes und einem sprecherspezifischen Anteil von Worten.

14. Verfahren nach Anspruch 13, wobei der vorgegebene Bestandteil des Wortschatzes festgelegte Worte enthält, die abhängig von einer objektivierten Gruppe von Sprechern ist.

15. Verfahren nach Anspruch 1, 5 oder 13, wobei im Profil zumindest ein Parameter für die Einstellung einer Syntaxerkennung bei der Wortfolgeerkennung als zweiten akustischen Erkennungsabschnitt (42) der Erkennungsumgebung (4) enthalten sind.

16. Verfahren nach Anspruch 1, wobei ein Profil zumindest einen Parameter zur Veränderung einer Dialogsteuerung (44) in der Erkennungsumgebung (4) besitzt.

17. Verfahren nach Anspruch 1, 5 oder 15, wobei ein Profil zumindest einen Parameter zur Beeinflussung einer semantischen Analyse (43) als einen dritten akustischen Erkennungsabschnitt in der Erkennungsumgebung (4) besitzt.

18. Verfahren nach Anspruch 1, wobei ein Profil (P_i) Parameter enthält für zumindest ein akustisches Modell (41) in Form von Merkmalen zur Anpassung des zumindest einen Modells an kehlkopfspezifische Merkmale des das Audiosignal erzeugenden Sprechers (Spracheigenschaft), zum Laden (8a) in einen akustischen Erkennungsabschnitt (41) der Erkennungsumgebung (4).

19. Verfahren nach Anspruch 1, wobei bei der Durchführung einer Sprachbearbeitung in der Erkennungsumgebung (4) gleichzeitig Schlüsselworte zur Steuerung des technischen Gerätes (52) und Suchworte zur Ermittlung vorgespeicherter Informationen, wie Telefon-Nummern, Benutzerkennungen, mit dem Audiosignal (s_a) verglichen, ohne eine Aufteilung der Sprachsignale in einen zeitlichen Abschnitt für die Erkennung von vorgespeicherter Information und einen zweiten zeitlichen Abschnitt für die Erkennung von Schlüsselworten.

20. Verfahren nach Anspruch 1, wobei die gespeicherten Profile (P_i) individuelle, insbesondere individualisierte Profile sind.

21. Verfahren nach Anspruch 2, wobei die Veränderung des Standard-Profils (P_x) zu einem individuellen Profil während der Dauer der Einbindung in die Erkennungsumgebung (4) erfolgt, wobei das gespeicherte Standard-Profil außerhalb der Erkennungsumgebung unverändert bleibt, aber die Individualisierung des geladenen Standard-Profils erfolgt:
durch Veränderung, insbesondere Reduzieren des vom Profil vorgegebenen Wortschatzes;
und/oder
durch Verändern der Parameter für die Eigenschaft einer optischen oder akustischen Dialogausgabe (44, 51);
und/oder
durch Verändern einer zur Erkennung zugelassenen Menge von Wortfolgen des vom Profil vorgegebenen Wortschatzes oder des reduzierten Wortschatzes.

22. Verfahren nach Anspruch 1, wobei mit einer Dialogausgabe (44, 51) als akustische oder optische Signalisierung ein Wechsel des in die Erkennungsumgebung (4) geladenen Profils vorbereitet oder veranlaßt wird.

23. Verfahren nach Anspruch 1, wobei die Wortfolgeerkennung (42) eingeschränkt wird, insbesondere durch Herabsetzen der verwendbaren Worte oder durch Einschränken der zur Erkennung zugelassenen Wortfolgen.

24. Verfahren nach Anspruch 23, wobei die Einschränkung nach Ausgabe eines Signals (51, 51a) zur optischen oder akustischen Signalisierung an einen Benutzer erfolgt.

25. Verfahren nach Anspruch 1, wobei beim Authentifizierungsversuch (2) über eine zweite Datenbank (31*) ein Benutzer einem individuellen Profil aus der ersten Datenbank (32) zugeordnet wird.

26. Verfahren nach Anspruch 25, wobei die zweite Datenbank (31*) Einträge besitzt, die geändert oder ergänzt werden, wenn ein neues individuelles Profil in der Erkennungsumgebung (4) erstellt und in der ersten Datenbank (32) abgespeichert wird.

27. Verfahren nach Anspruch 1 oder 18, wobei zur Verbesserung der Erkennung in der Erkennungsumgebung (4) in einem akustischen Modell (41) eine Anpassung an eine Art der Übertragung oder eine Art der Aufnahme des aktuellen Audiosignals (s_a) erfolgt.

28. Verfahren nach Anspruch 1, wobei die Wortfolgeerkennung (42) nach einem erkannten Wort für das nächste zu erkennende Wort eine Begrenzung der verfügbaren Worte vornimmt, die durch das Profil zugeordnet oder vorgegeben werden.

29. Verfahren nach Anspruch 28, wobei das Audiosignal in einer erkannten Kette von Worten gespeichert wird.

30. Verfahren zum sprachgesteuerten Betreiben eines Gerätes (**52**), wobei

in einer ersten Linie (F_A) gemäß den Merkmalen (b) und (c) des Anspruchs 1 ein erstes Profil (**31A**) entsprechend dem Ausgangssignal einer gemäß Merkmal (a) des Anspruchs 1 aufgebauten und arbeitsfähigen Authentifizierung (**2**) ausgewählt und einer ersten Erkennungsumgebung (**4A**) zugeordnet wird; und

in einer zweiten Linie (F_B) auch entsprechend den Merkmalen (b) und (c), aber gesondert von der vorhergehenden Linie (F_A) eine Auswahl und eine Einbindung eines anderen Profils (**31B**) in einer zweiten Erkennungsumgebung (**4B**) erfolgt, auch abhängig von dem Ausgangssignal (a*) der Authentifizierung (**2**);

eine Entscheidungseinrichtung (**6**) vorgesehen ist, die Ausgangssignale (A, B) der Erkennungsumgebungen (**4A, 4B**) der beiden Linien (F_A, F_B) bewertet, um eine der beiden auszuwählen und dem zu steuernden Gerät (**52**) Steuersignale zuzuführen, die der Bedeutung des ausgewählten Signals entweder der ersten oder der zweiten Erkennungslinie entsprechen.

31. Verfahren nach Anspruch 1 oder 30, wobei jede Erkennungsumgebung aus dem Audiosignal (s_a) entsprechend dem eingebundenen oder geladenen Profil einen eigenständigen Bedeutungsgehalt ermittelt, der einem Steuersignal entspricht, das dem zu steuernden Gerät (**52**) über einen Eingang (**52a**) zuführbar ist.

32. Verfahren nach Anspruch 30, wobei eine dritte eigenständige Linie (F_X) vorgesehen ist, die entsprechend der ersten und zweiten Linie ausgebildet ist und deren Ausgangssignal (C) einem Bedeutungsgehalt zumindest eines Abschnitts des Audiosignals entspricht, um der Entscheidungseinrichtung (**6**) zugeführt zu werden, wobei die Entscheidungseinrichtung (**6**) denjenigen Bedeutungsgehalt aus den drei zugeführten Bedeutungen (A, B, C) auswählt, der in der Mehrzahl ist, um ihn an das Gerät (**52**) über dessen Eingang (**52a**) zu übertragen.

33. Verfahren nach Anspruch 30, mit einem Verfahren nach einem der Ansprüche 1 bis 29.

34. Einrichtung mit Einzelfunktionen, arbeitsfähig nach einem der vorgenannten Verfahrensansprüche.

35. Verfahren zum Betreiben einer sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technisches Gerätes (**52**), wie eine Telefonvermittlung oder eine Telefonanlage; wobei ein Audiosignal (s_a) aus zumindest einem von einem Sprecher abgegebenen Wort, insbesondere mehreren aufeinanderfolgenden Worten, welche einen Bedeutungsgehalt besitzen und insoweit erkannt werden sollen, einem ersten Signaleingang (**10**) einer Authentifizierungseinrichtung (**2**) zugeführt werden; wobei

(a) eine Sprechererkennung aufgrund eines Authentifizierungsversuchs (**2**) erfolgt, insbesondere durch einen zeitlichen Abschnitt des Audiosignals (s_a), durch Erkennen einer Selbstauthentifizierung mittels aktiven Zutuns, oder durch Auswerten von Sekundärindizien (**12**), wie eine bekannte Telefon-Nummer, ein Nebenstellen-Anschluß, um einen Sprecher als Individuum oder eine objektivierte Gruppe von Sprechern festzulegen, der der Sprecher durch objektivierte Kriterien des auf ihn zurückzuführenden Audiosignals zuzuordnen ist, und ein entsprechendes Ausgangssignal (a, a*) abzugeben;

(b) ein zum festgelegten Sprecher oder der objektivierten Gruppe korrespondierendes Profil (**33**) aus einer Vielzahl gespeicherter Profile (32, P_i) ausgewählt (**2, 31**) wird, gestützt auf das Ausgangssignal (a, a*) der Authentifizierung;

(c) das ausgewählte Profil (**33**) in eine Erkennungsumgebung (**4**) eingebunden oder geladen wird, um die Erkennungsumgebung auf den festgelegten Sprecher bzw. die objektivierte Gruppe hin anzupassen; wobei jedes der gespeicherten Profile (P_i) und das eingebundene oder geladene Profit (**33**) Parameter enthält, zur Beeinflussung zumindest einer in der Erkennungsumgebung (**4**) vorgesehenen Wortfolgeerkennung (**42**);

(d) in der Erkennungsumgebung (**4**) zumindest eine Wortfolgeerkennung (**42**) als lexikalische Analyse vorgesehen ist, bei der aus zumindest zwei Worten des akustischen Signals eine Bedeutung der Wortfolge so erkannt wird, daß nach Erkennen eines ersten Wortes für das nächstfolgende Wort nur eine begrenzte Anzahl von Worten des durch das Profil zugeordneten Wortschatzes zugelassen wird.

Hierzu 3 Seite(n) Zeichnungen

Nummer: DE 100 50 808 A1
Int. Cl.[7]: G 10 L 15/22
Offenlegungstag: 16. Mai 2002

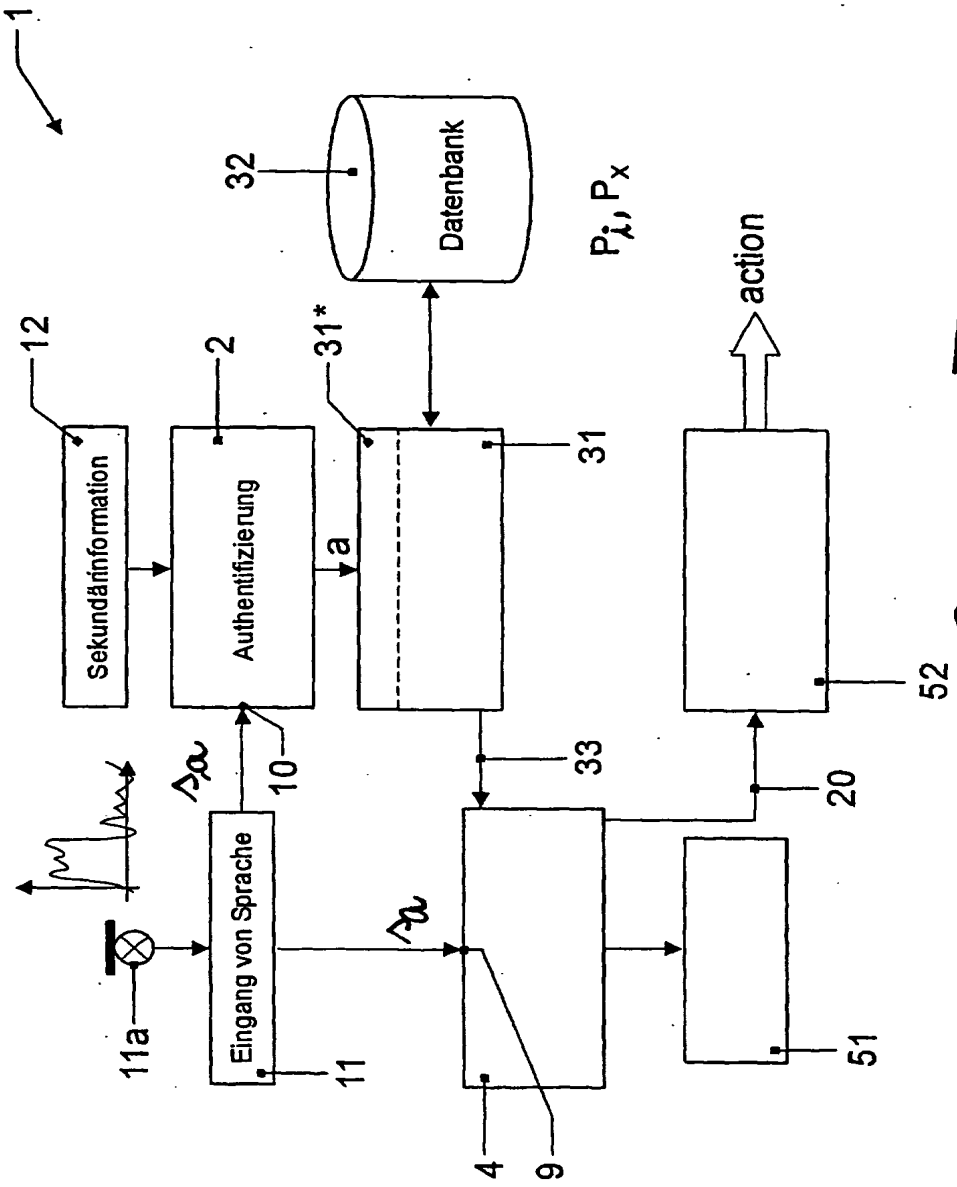


Fig. 1

P_i ; i=1...n
P_x

- Leerseite -

Nummer: DE 100 50 808 A1
Int. Cl.[7]: G 10 L 15/22
Offenlegungstag: 16. Mai 2002

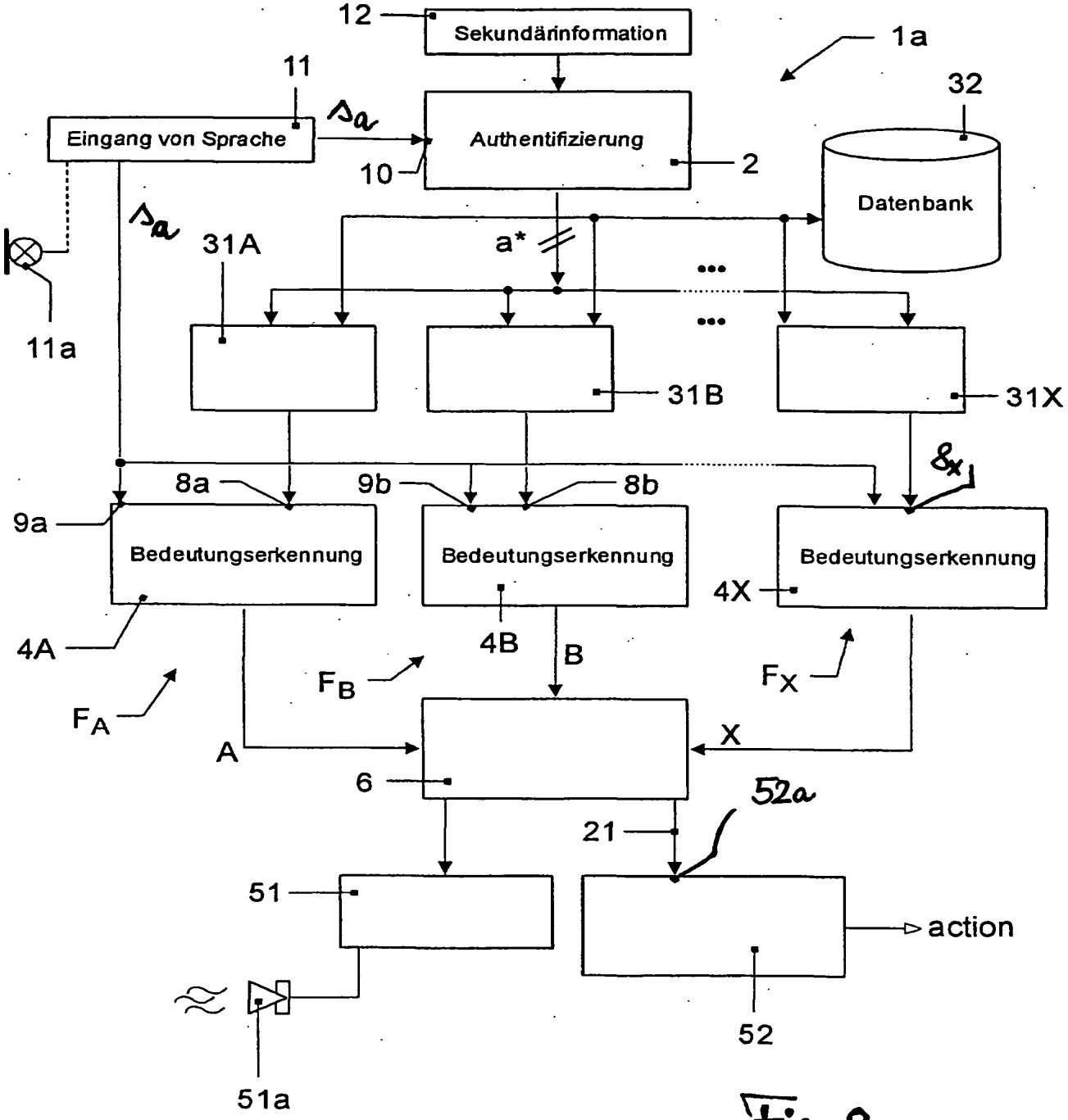
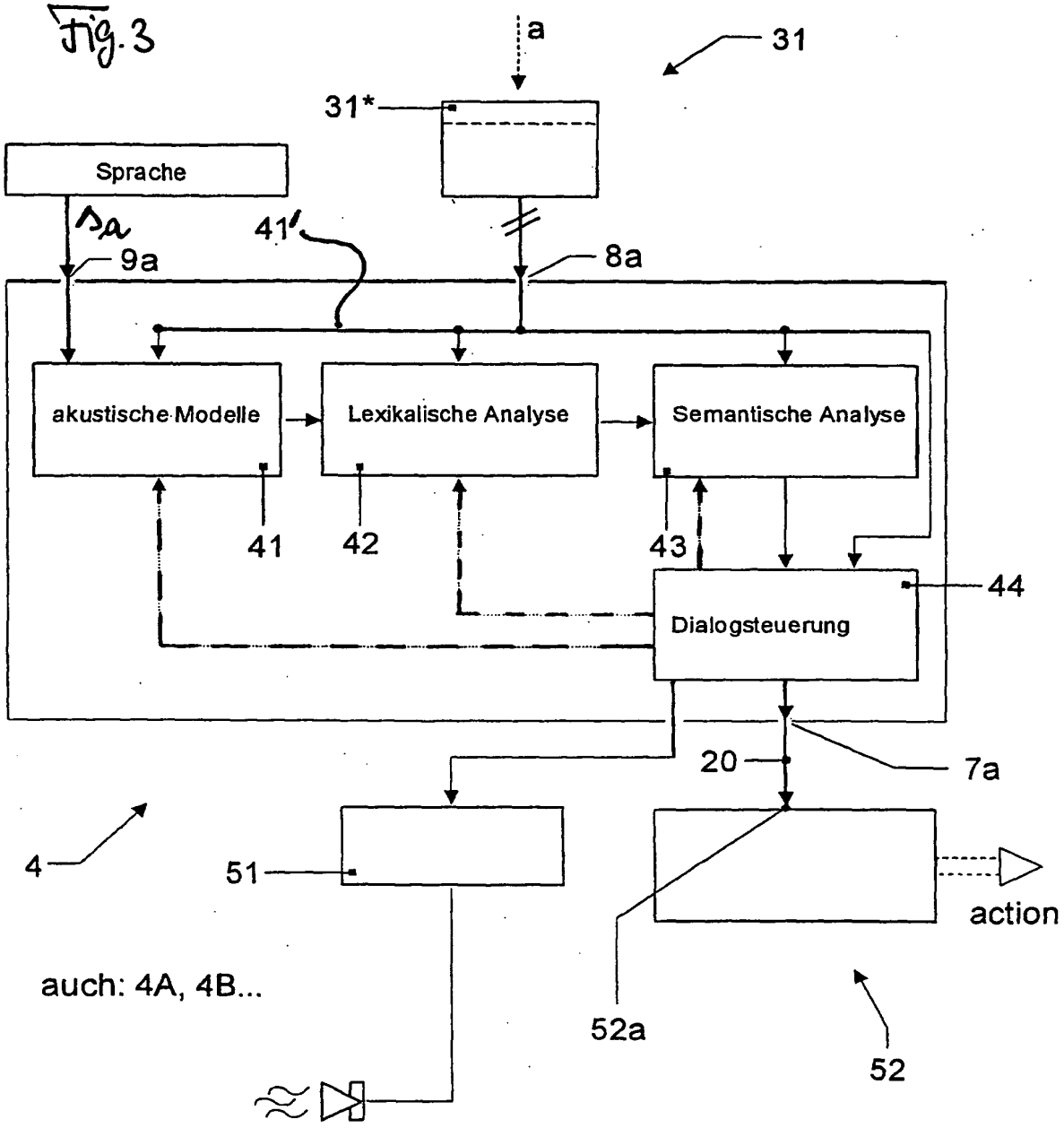


Fig. 2

Nummer:
Int. Cl.[7]:
Offenlegungstag:

DE 100 50 808 A1
G 10 L 15/22
16. Mai 2002

Fig. 3



(12) NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT) VERÖFFENTLICHTE INTERNATIONALE ANMELDUNG

(19) Weltorganisation für geistiges Eigentum
Internationales Büro

(43) Internationales Veröffentlichungsdatum
18. April 2002 (18.04.2002)

PCT

(10) Internationale Veröffentlichungsnummer
WO 02/31813 A1

(51) Internationale Patentklassifikation[7]: G10L 15/06, 17/00

(21) Internationales Aktenzeichen: PCT/DE01/03925

(22) Internationales Anmeldedatum: 12. Oktober 2001 (12.10.2001)

(25) Einreichungssprache: Deutsch

(26) Veröffentlichungssprache: Deutsch

(30) Angaben zur Priorität:
100 50 808.1 13. Oktober 2000 (13.10.2000) DE

(71) Anmelder *(für alle Bestimmungsstaaten mit Ausnahme von US)*: **VOICECOM AG** [DE/DE]; Südwestpark 48, 90449 Nürnberg (DE).

(72) Erfinder; und

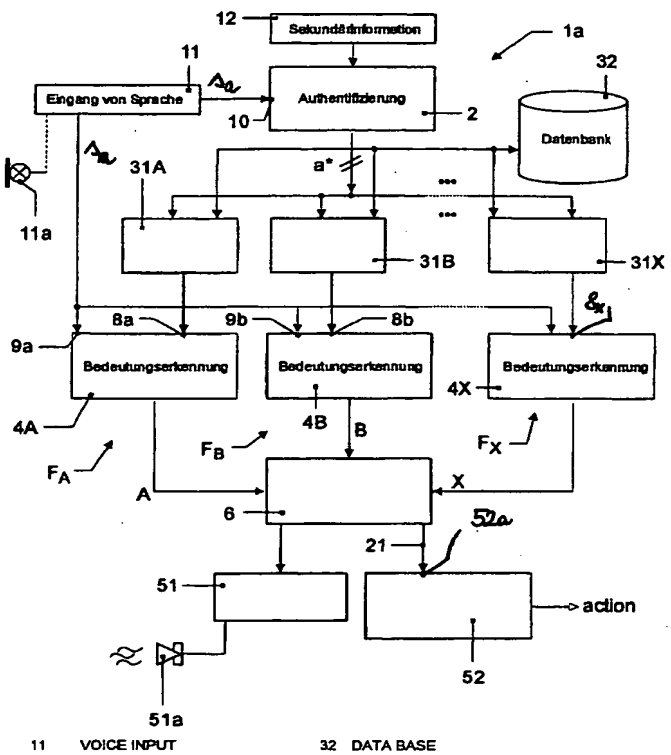
(75) Erfinder/Anmelder *(nur für US)*: **SCHIMMER, Klaus** [DE/DE]; Kranichweg 20, 90513 Zirndorf (DE). **PLANKENSTEINER, Peter** [DE/DE]; Genglerstrasse 13, 91054 Erlangen (DE). **HARBECK, Stefan** [DE/DE]; Drausnickstrasse 114, 91052 Erlangen (DE).

(74) Anwälte: **LEONHARD OLGEMOELLER FRICKE** usw.; Postfach 10 09 57, 80083 München (DE).

[Fortsetzung auf der nächsten Seite]

(54) Title: VOICE-DRIVEN DEVICE CONTROL WITH AN OPTIMISATION FOR A USER

(54) Bezeichnung: SPRACHGEFÜHRTE GERÄTESTEUERUNG MIT EINER OPTIMIERUNG FÜR EINEN BENUTZER



11 VOICE INPUT
12 SECONDARY INFORMATION
2 AUTHENTIFICATION
32 DATA BASE
4 RECOGNITION OF MEANING

(57) **Abstract:** The invention relates to a method for preparing, operating or adapting a voice-driven control device for operating a technical device (52). An audio signal (sa) consisting of at least one word uttered by a speaker is fed to a first signal input (10) of an authentication device (2). A speaker recognition operation is carried out on the basis of an authentication attempt (2) in order to establish a speaker as an individual or as an objectivated group of speakers to which the speaker is to be assigned through objectivated criteria of the audio signal to be traced back to the speaker, and in order to deliver a corresponding output signal (a, a*). A profile (33) corresponding to the established speaker or the objectivated group is selected (2,31) from a number of stored profiles (32,Pi) with the aid of the output signal (a, a*) from the authentication, and the selected profile (33) is integrated or loaded into a recognition environment (4) in order to adapt the recognition environment to the established speaker or the objectivated group. Each of the stored profiles (Pi) and the integrated or loaded profile (33) contain parameters for influencing at least one recognition of a word sequence (42) provided in the recognition environment (4).

(57) **Zusammenfassung:** Die Erfindung bezieht sich auf ein Verfahren zum Vorbereiten, Betreiben oder Anpassen einer sprachgesteuerten Steuerungseinrichtung zur Bedienung eines technisches Gerätes (52), wobei ein Audiosignal (sa) aus zumindest einem von einem Sprecher abgegebenen

[Fortsetzung auf der nächsten Seite]

WO 02/31813 A1